Appendix

Detailed Methods

Data Collection

We retrieved all available hepatitis C virus sequence records from the National Center for Biotechnology Information (NCBI) Genbank nucleotide database using the search query "hepatitis C[organism]". The search results, comprising 160,556 records, were parsed using the BioPython SeqIO module[1]. Any records with `Patent' in the description field or lacking a source modifier field (attribute information associated with the sequence) were excluded (*n*=114,696).

From the remaining records, sample collection dates were extracted from the source modifier field if available (*n* = 45,316 records), and the corresponding dated sequences were written to a FASTA-formatted file. The remaining 544 sequence records were manually screened for any collection date information in the definition or description fields, recovering an additional 23 sequences. Specific gene regions were extracted from these data by using a pairwise alignment algorithm[2,3] to align each sequence against the respective sequence intervals from the H77 reference genome and extracting the overlapping region. Sequences were then screened to remove any that may have originated from the same patient or from epidemiologically linked patients first by calculating a inter vs. intra patient phylogenetic distance cutoff where all but one sequences falling within the intra patient range were excluded from analysis thus only one sequence per patient is included, and second by checking the reference associated with the sequence from the reference field on GenBank. Each gene-specific dataset was manually inspected and clipped to extract the broadest interval with the greatest overlap

across sequences. We selected genomic regions for phylodynamic analyses based on previous whole genome studies of HCV genotype one which revealed E2, NS2, and NS5B as the most informative HCV gene regions for phylodynamic studies[4]. Additionally, we analyzed the E1 and NS4B regions[4].

Data Curation

To screen for genotype 1a sequences, genotype and gene specific reference sequences were collected from the Los Alamos National Laboratory HCV sequence repository. Sequences in each dataset were aligned using MAFFT v7·154b[5] and manually inspected using HYPHY v2·22[3]. Phylogenetic trees for each complete dataset were inferred in an approximate maximum likelihood modeling framework as implemented in FastTree2[6]. To root the resulting phylogenetic trees under the assumption of a molecular clock, we selected roots that minimized root mean square error between root-to-tip distances and sampling dates as implemented in the RTT function in the R package APE[7]. Phylogenetic trees were then visualized and manually inspected using FigTree (http://tree.bio.ed.ac.uk/software/figtree/) to isolate the clade comprising all sequences annotated as genotype 1a. This procedure identified 13 additional sequences that were incorrectly annotated with a different genotype. Following curation our HCV gene-specific datasets comprised 917 sequences and 4,449 bases (E1: *n*=252, 576 bp; E2: *n*=196, 1,089 bp; NS2: *n*=182, 651 bp, NS4B: *n*=139 NS4B, 783 bp; NS5B *n*=148, 1400 bp). Sequences were annotated with accession number, collection date, and geographic location of origin. 88 sets of sequences originated from full-length genome sequences. Accession numbers for all public domain sequences used can be found in Appendix Table

S1. Sampling times of included sequences ranged from 1977-2011, Appendix Figure 2. Approximate maximum likelihood phylogenetic trees for each gene region are displayed in Appendix Figure 3.

Effective Number of HCV Infections

We reconstructed the dynamics of the North American HCV epidemic over time using Bayesian skyline plots[8]. The y-axis in a Bayesian skyline plot represents N_e multiplied by *tau*, where N_e is the effective population size and tau is the generation time. In this context the y-axis (N_e * *tau*) is interpreted as the number of infected individuals who go on to infect additional individuals (effective number of infections[9]) on a $\log_{10}$ scale, and the x-axis represents time in years. To determine if choice of priors on both the model by which sequences accrue changes (substitution models) or the model by which the expected number of changes in a sequence relates to time (clock models) biased results, we performed two runs for each gene with four different substitution models (General Time Reversible (GTR) and Tamura-Nei 93 (TN93) with and without gamma distributed rate variation) and three different clock models (strict, random local, and relaxed). To determine the best fitting performed model comparison in the program Tracer v1·5 (http://tree.bio.ed.ac.uk) using Bayes Factors[10]. Optimal run conditions were determined to be a General Time Reversible substitution model with among site rate heterogeneity distributed according to a gamma distribution under a relaxed molecular clock. Replicate chains of Markov Chain Monte Carlo (MCMC) sampling implemented in BEAST v·1·8·0[11] were run a minimum of four times per gene for at least 5×10^8 generations per

run. The convergence of each run was confirmed through evaluation of parameter traces and effective sample size (ESS) values greater than 200 via the program Tracer v 1·5.

Appendix Table 1. Accession numbers used in analyses for each analyzed HCV genomic region. All sequences are from North America.

E1	E2	NS2	NS4B	NS5B
AB079078	DQ430811.1	DQ430811.1	DQ430812.1	DQ430812.1
AB079081	DQ430813.1	DQ430813.1	DQ430814.1	DQ430813.1
AB079088	EU155248.2	EU155248.2	EU155248.2	EU155248.2
DQ430812.1	EU155287.2	EU155287.2	EU155287.2	EU155287.2
DQ430813.1	EU155288.2	EU155288.2	EU155288.2	EU155288.2
EU155248.2	EU155309.2	EU155309.2	EU155299.2	EU155309.2
EU155287.2	EU155310.2	EU155310.2	EU155310.2	EU155310.2
EU155288.2	EU155338.2	EU155338.2	EU155338.2	EU155338.2
EU155309.2	EU155339.2	EU155339.2	EU155339.2	EU155339.2
EU155310.2	EU155340.2	EU155340.2	EU155340.2	EU155340.2
EU155338.2	EU155341.2	EU155341.2	EU155341.2	EU155341.2
EU155339.2	EU155342.2	EU155342.2	EU155342.2	EU155342.2
EU155340.2	EU255942.1	EU255942.1	EU255942.1	EU255942.1
EU155341.2	EU255963.1	EU255967.1	EU255966.1	EU255967.1
EU155342.2	EU255967.1	EU255975.1	EU255975.1	EU255975.1
EU234065.2	EU255975.1	EU255988.1	EU255988.1	EU255976.1
EU255942.1	EU255988.1	EU862827.1	EU862827.1	EU255988.1
EU255967.1	EU862827.1	FJ390399.1	FJ024281.1	EU862827.1
EU255975.1	FJ024282.1	FJ410172.1	FJ024282.1	FJ205868.1
EU255988.1	FJ390394.1	GQ925446.1	FJ410172.1	FJ390394.1
FJ024281.1	FJ688411.1	HM000514.1	JF779679.1	FJ390399.1
FJ024282.1	GQ379232.1	HM000515.1	JX463525.1	JF424738.1
FJ207570.1	HM000514.1	HM000516.1	JX463526.1	JF424744.1
FJ207577.1	HM000515.1	HM000517.1	JX463527.1	JF424745.1
FJ207596.1	HM000516.1	JQ801756.1	JX463528.1	JF424747.1
FJ207598.1	HM000517.1	JQ801812.1	JX463529.1	JF424748.1
FJ207606.1	JQ063473.1	JQ801885.1	JX463530.1	JF424750.1
FJ207607.1	JQ268595.1	JQ801886.1	JX463531.1	JF779679.1
FJ207621.1	JQ801756.1	JQ801920.1	JX463533.1	JX463525.1
FJ207622.1	JQ801812.1	JQ801975.1	JX463534.1	JX463526.1
FJ207638.1	JQ801853.1	JQ801978.1	JX463535.1	JX463527.1
FJ207639.1	JQ801885.1	JQ802051.1	JX463536.1	JX463528.1
FJ207658.1	JQ801886.1	JQ802052.1	JX463537.1	JX463529.1
FJ207659.1	JQ801921.1	JQ802117.1	JX463538.1	JX463530.1
FJ207680.1	JQ801977.1	JQ802118.1	JX463539.1	JX463533.1
FJ207682.1	JQ801978.1	JQ802322.1	JX463540.1	JX463534.1
FJ207702.1	JQ802051.1	JQ802344.1	JX463541.1	JX463535.1
FJ207747.1	JQ802052.1	JQ802423.1	JX463542.1	JX463536.1

FJ207748.1	JQ802113.1	JQ802424.1	JX463543.1	JX463537.1
FJ207765.1	JQ802143.1	JQ802469.1	JX463544.1	JX463538.1
FJ207786.1	JQ802289.1	JQ802480.1	JX463545.1	JX463539.1
FJ207787.1	JQ802366.1	JQ802513.1	JX463546.1	JX463540.1
FJ207807.1	JQ802376.1	JQ802544.1	JX463547.1	JX463541.1
FJ207878.1	JQ802439.1	JQ802657.1	JX463548.1	JX463542.1
FJ207879.1	JQ802451.1	JQ802670.1	JX463549.1	JX463543.1
FJ207896.1	JQ802507.1	JQ802750.1	JX463550.1	JX463544.1
FJ207897.1	JQ802512.1	JQ802753.1	JX463551.1	JX463545.1
FJ207911.1	JQ802600.1	JQ802871.1	JX463552.1	JX463546.1
FJ207912.1	JQ802624.1	JQ802874.1	JX463553.1	JX463547.1
FJ390394.1	JQ802705.1	JQ802994.1	JX463554.1	JX463548.1
FJ390399.1	JQ802714.1	JQ803126.1	JX463555.1	JX463549.1
GQ379232.1	JQ802800.1	JQ803258.1	JX463556.1	JX463550.1
GU580965.1	JQ802824.1	JQ803262.1	JX463557.1	JX463551.1
GU580968.1	JQ802946.1	JQ803382.1	JX463558.1	JX463552.1
GU580970.1	JQ802948.1	JQ803383.1	JX463559.1	JX463553.1
GU580982.1	JQ803174.1	JQ804158.1	JX463560.1	JX463554.1
HM000514.1	JQ803185.1	JQ804216.1	JX463561.1	JX463555.1
HM000515.1	JQ803344.1	JQ804250.1	JX463563.1	JX463556.1
HM000516.1	JQ803345.1	JQ804258.1	JX463564.1	JX463557.1
HM000517.1	JQ803398.1	JQ804361.1	JX463565.1	JX463558.1
HM350733.1	JQ803688.1	JQ804378.1	JX463566.1	JX463559.1
HM350801.1	JQ803759.1	JQ804512.1	JX463567.1	JX463560.1
HM350803.1	JQ803813.1	JX178303.1	JX463568.1	JX463561.1
HM350887.1	JQ803880.1	JX178372.1	JX463569.1	JX463563.1
HM350889.1	JQ803888.1	JX178377.1	JX463570.1	JX463564.1
HM350950.1	JQ803948.1	JX178440.1	JX463571.1	JX463565.1
HM350961.1	JQ804028.1	JX463525.1	JX463572.1	JX463566.1
HM351021.1	JQ804131.1	JX463526.1	JX463573.1	JX463567.1
HM351046.1	JX178319.1	JX463527.1	JX463574.1	JX463568.1
HM351084.1	JX178335.1	JX463528.1	JX463575.1	JX463569.1
HM351090.1	JX178414.1	JX463529.1	JX463576.1	JX463570.1
HM351159.1	JX463525.1	JX463530.1	JX463577.1	JX463571.1
HM351161.1	JX463526.1	JX463531.1	JX463578.1	JX463572.1
HM351308.1	JX463527.1	JX463533.1	JX463579.1	JX463573.1
HM351328.1	JX463528.1	JX463534.1	JX463580.1	JX463574.1
HM351428.1	JX463529.1	JX463535.1	JX463581.1	JX463575.1
HM351438.1	JX463530.1	JX463536.1	JX463582.1	JX463576.1
HM351547.1	JX463532.1	JX463537.1	JX463583.1	JX463577.1
HM351557.1	JX463533.1	JX463538.1	JX463584.1	JX463578.1
HM351649.1	JX463534.1	JX463539.1	JX463585.1	JX463579.1
HM351686.1	JX463535.1	JX463540.1	JX463586.1	JX463580.1

HM351776.1	JX463536.1	JX463541.1	JX463587.1	JX463581.1
HM351777.1	JX463537.1	JX463542.1	JX463588.1	JX463582.1
HM352026.1	JX463538.1	JX463543.1	JX463589.1	JX463583.1
HM352037.1	JX463539.1	JX463544.1	JX463590.1	JX463584.1
HM352096.1	JX463540.1	JX463545.1	JX463591.1	JX463585.1
HM352098.1	JX463541.1	JX463546.1	JX463592.1	JX463586.1
HM352198.1	JX463542.1	JX463547.1	JX463593.1	JX463587.1
HM352205.1	JX463543.1	JX463548.1	JX463594.1	JX463588.1
HM352311.1	JX463544.1	JX463549.1	JX463595.1	JX463589.1
JQ063480.1	JX463545.1	JX463550.1	JX463596.1	JX463590.1
JQ268595.1	JX463546.1	JX463551.1	JX463597.1	JX463591.1
JQ801756.1	JX463547.1	JX463552.1	JX463598.1	JX463592.1
JQ801811.1	JX463548.1	JX463553.1	JX463599.1	JX463593.1
JQ801853.1	JX463549.1	JX463554.1	JX463600.1	JX463594.1
JQ801885.1	JX463550.1	JX463555.1	JX463601.1	JX463595.1
JQ801892.1	JX463551.1	JX463556.1	JX463602.1	JX463596.1
JQ801920.1	JX463552.1	JX463557.1	JX463603.1	JX463597.1
JQ801975.1	JX463553.1	JX463558.1	JX463604.1	JX463598.1
JQ801979.1	JX463554.1	JX463559.1	JX463605.1	JX463599.1
JQ802050.1	JX463555.1	JX463560.1	JX463606.1	JX463600.1
JQ802053.1	JX463556.1	JX463561.1	JX463607.1	JX463601.1
JQ802113.1	JX463557.1	JX463563.1	JX463608.1	JX463602.1
JQ802143.1	JX463558.1	JX463564.1	JX463609.1	JX463603.1
JQ802322.1	JX463559.1	JX463565.1	JX463610.1	JX463604.1
JQ802345.1	JX463560.1	JX463566.1	JX463611.1	JX463605.1
JQ802422.1	JX463561.1	JX463567.1	JX463612.1	JX463606.1
JQ802440.1	JX463563.1	JX463568.1	JX463613.1	JX463607.1
JQ802460.1	JX463564.1	JX463569.1	JX463614.1	JX463608.1
JQ802477.1	JX463565.1	JX463570.1	JX463615.1	JX463609.1
JQ802603.1	JX463566.1	JX463571.1	JX463617.1	JX463610.1
JQ802616.1	JX463567.1	JX463572.1	JX463618.1	JX463611.1
JQ802704.1	JX463568.1	JX463573.1	JX463619.1	JX463612.1
JQ802718.1	JX463569.1	JX463574.1	JX463620.1	JX463613.1
JQ802802.1	JX463570.1	JX463575.1	JX463621.1	JX463614.1
JQ802814.1	JX463571.1	JX463576.1	JX463622.1	JX463615.1
JQ802946.1	JX463572.1	JX463577.1	JX463623.1	JX463616.1
JQ802947.1	JX463573.1	JX463578.1	JX463624.1	JX463617.1
JQ803149.1	JX463574.1	JX463579.1	JX463625.1	JX463618.1
JQ803182.1	JX463575.1	JX463580.1	JX463626.1	JX463619.1
JQ803337.1	JX463576.1	JX463581.1	JX463627.1	JX463620.1
JQ803350.1	JX463577.1	JX463582.1	JX463628.1	JX463621.1
JQ803398.1	JX463578.1	JX463583.1	JX463629.1	JX463622.1
JQ803658.1	JX463579.1	JX463584.1	JX463630.1	JX463623.1

JQ803759.1	JX463580.1	JX463585.1	JX463631.1	JX463624.1
JQ803827.1	JX463581.1	JX463586.1	JX463632.1	JX463625.1
JQ803893.1	JX463582.1	JX463587.1	JX463634.1	JX463626.1
JQ804028.1	JX463583.1	JX463588.1	JX463635.1	JX463627.1
JX178314.1	JX463584.1	JX463589.1	JX463636.1	JX463628.1
JX178362.1	JX463585.1	JX463590.1	JX463637.1	JX463629.1
JX178399.1	JX463586.1	JX463591.1	JX463638.1	JX463630.1
JX178443.1	JX463587.1	JX463592.1	JX463639.1	JX463631.1
JX463525.1	JX463588.1	JX463593.1	JX463640.1	JX463632.1
JX463529.1	JX463589.1	JX463594.1	JX463641.1	JX463633.1
JX463530.1	JX463590.1	JX463595.1	JX472008.1	JX463634.1
JX463531.1	JX463591.1	JX463596.1	JX472009.1	JX463635.1
JX463533.1	JX463592.1	JX463597.1	KM102760.1	JX463636.1
JX463534.1	JX463593.1	JX463598.1	KM102762.1	JX463637.1
JX463536.1	JX463594.1	JX463599.1	KM102763.1	JX463638.1
JX463537.1	JX463595.1	JX463600.1		JX463639.1
JX463538.1	JX463596.1	JX463601.1		JX463640.1
JX463539.1	JX463597.1	JX463602.1		JX463641.1
JX463540.1	JX463598.1	JX463603.1		JX472005.1
JX463541.1	JX463599.1	JX463604.1		JX472009.1
JX463542.1	JX463600.1	JX463605.1		KM102760.1
JX463543.1	JX463601.1	JX463606.1		KM102761.1
JX463544.1	JX463602.1	JX463607.1		KM102762.1
JX463545.1	JX463603.1	JX463608.1		KM102763.1
JX463546.1	JX463604.1	JX463609.1		
JX463547.1	JX463605.1	JX463610.1		
JX463549.1	JX463607.1	JX463611.1		
JX463550.1	JX463608.1	JX463612.1		
JX463551.1	JX463609.1	JX463613.1		
JX463552.1	JX463610.1	JX463614.1		
JX463553.1	JX463611.1	JX463615.1		
JX463554.1	JX463612.1	JX463616.1		
JX463555.1	JX463613.1	JX463617.1		
JX463556.1	JX463614.1	JX463618.1		
JX463557.1	JX463615.1	JX463619.1		
JX463558.1	JX463616.1	JX463620.1		
JX463559.1	JX463617.1	JX463621.1		
JX463561.1	JX463618.1	JX463622.1		
JX463562.1	JX463619.1	JX463623.1		
JX463564.1	JX463620.1	JX463624.1		
JX463565.1	JX463621.1	JX463625.1		
JX463566.1	JX463622.1	JX463626.1		
JX463567.1	JX463623.1	JX463627.1		

JX463568.1
JX463569.1
JX463570.1
JX463571.1
JX463572.1
JX463573.1
JX463574.1
JX463575.1
JX463576.1
JX463577.1
JX463578.1
JX463579.1
JX463580.1
JX463581.1
JX463582.1
JX463584.1
JX463585.1
JX463586.1
JX463587.1
JX463588.1
JX463589.1
JX463590.1
JX463591.1
JX463592.1
JX463593.1
JX463594.1
JX463595.1
JX463596.1
JX463597.1
JX463599.1
JX463600.1
JX463601.1
JX463602.1
JX463603.1
JX463604.1
JX463606.1
JX463607.1
JX463608.1
JX463609.1
JX463610.1
JX463611.1
JX463613.1
JX463614.1
JX463624.1
JX463625.1
JX463626.1
JX463627.1
JX463628.1
JX463629.1
JX463630.1
JX463631.1
JX463632.1
JX463633.1
JX463634.1
JX463635.1
JX463636.1
JX463637.1
JX463638.1
JX463639.1
JX463640.1
JX463641.1
JX472005.1
JX472009.1
KC614825.1
KC614830.1
KC614869.1
KC614875.1
KC614920.1
KC614926.1
KC615031.1
KC615232.1
KJ210360.1
JX463628.1
JX463629.1
JX463630.1
JX463631.1
JX463632.1
JX463633.1
JX463634.1
JX463635.1
JX463637.1
JX463638.1
JX463639.1
JX463640.1
JX463641.1
JX472005.1
JX472009.1

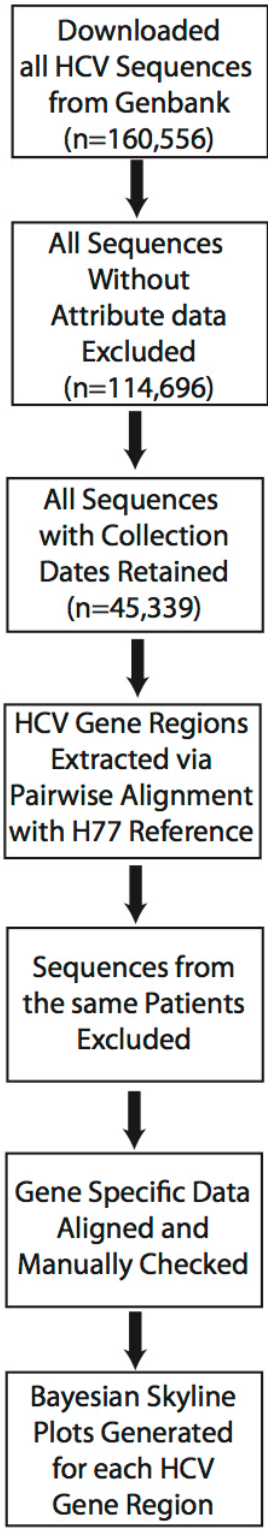
JX463615.1
JX463617.1
JX463618.1
JX463619.1
JX463620.1
JX463621.1
JX463622.1
JX463623.1
JX463624.1
JX463625.1
JX463626.1
JX463627.1
JX463628.1
JX463629.1
JX463630.1
JX463631.1
JX463632.1
JX463633.1
JX463634.1
JX463635.1
JX463636.1
JX463637.1
JX463639.1
JX463640.1
JX472008.1
JX472009.1
KC562896.1
KC614816.1
KC614829.1
KC614873.1
KC614876.1
KC614923.1
KC614926.1
KC615032.1
KC615033.1
KC615088.1
KC615089.1
KC615157.1
KC615163.1
KC615204.1
KC615217.1
KJ210360.1

1

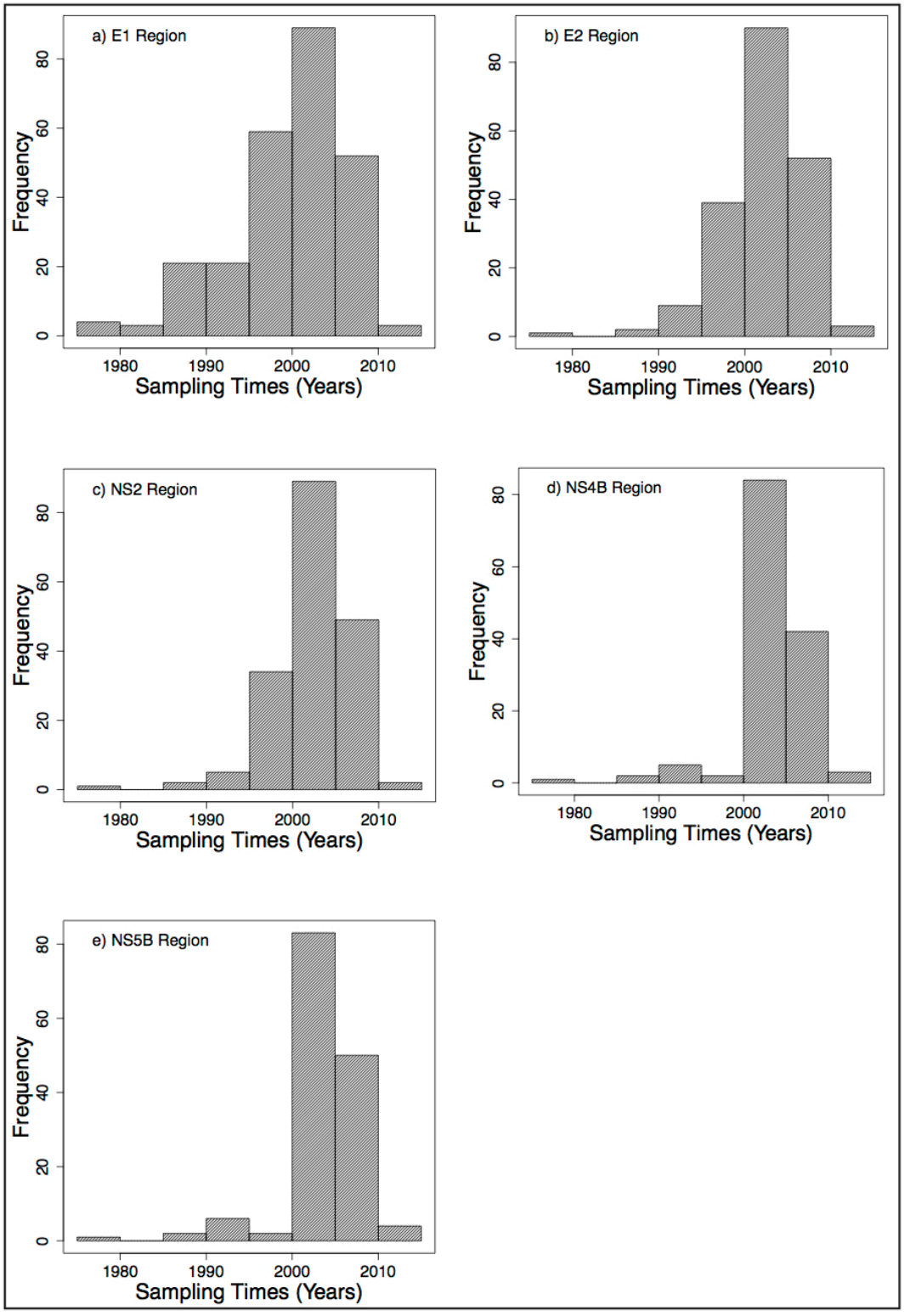
Appendix Table 2. Comparison of the number of sequences by 5-year interval and by HCV gene region included in this study compared to that of a similar previous study[4].

Time Interval	Previous Study NS5B	This Study E1	This Study E2	This Study NS2	This Study NS4B	This Study NS5B
<1980	1	4	1	1	1	1
1980 - 1984	0	2	0	0	0	0
1985 - 1989	1	20	1	1	1	1
1990 - 1994	4	18	9	5	5	6
1995 - 1999	2	61	39	34	3	3
2000 - 2004	44	77	75	73	69	68
2005 - 2009	35	66	67	66	56	64
2010 - 2011	0	4	4	2	4	5
Total	**87**	**252**	**196**	**182**	**139**	**148**

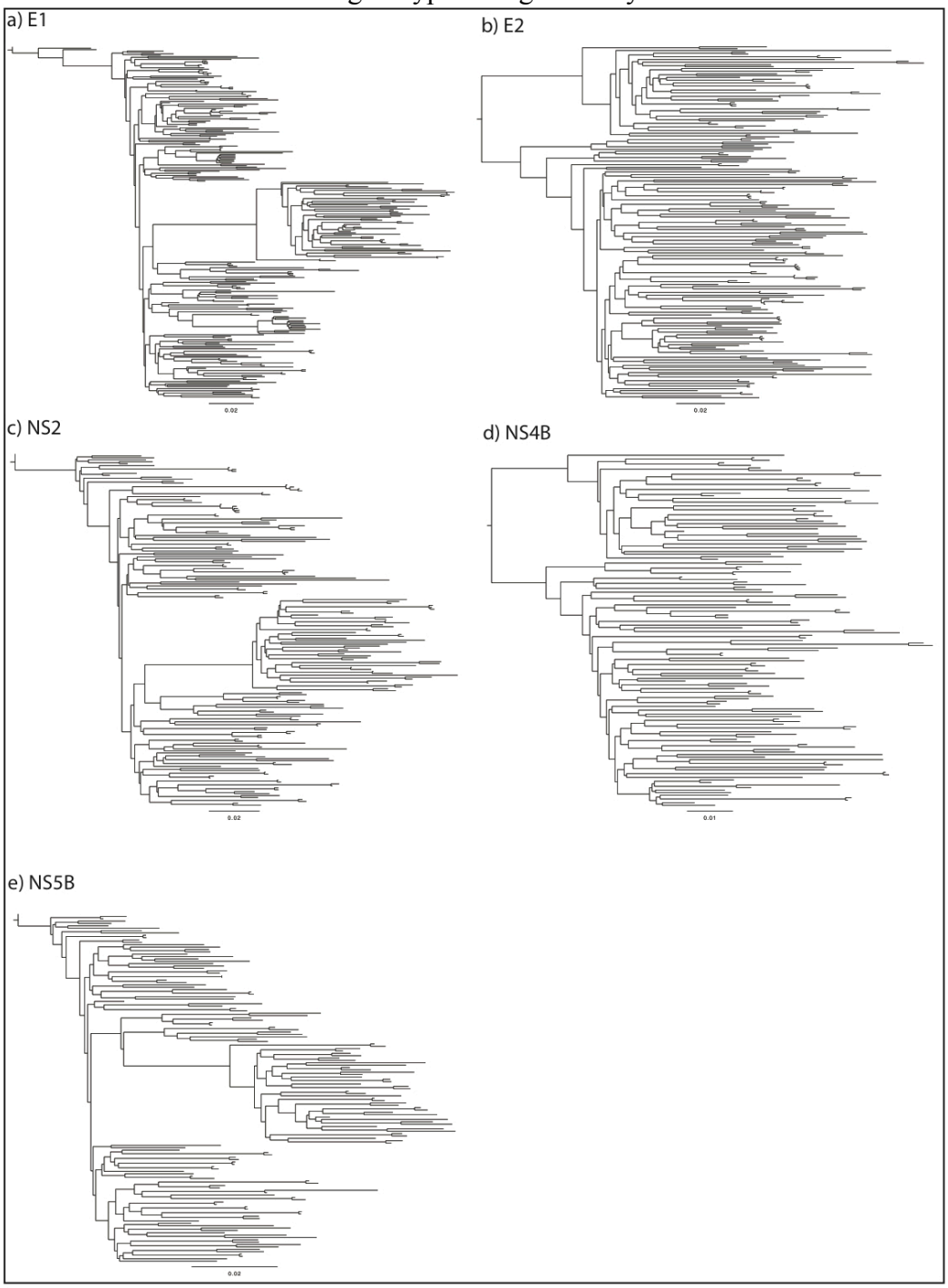
Appendix Figure 1. Flow chart detailing the steps involved in data collection, curation and analysis. Sample sizes of sequences are included in brackets.



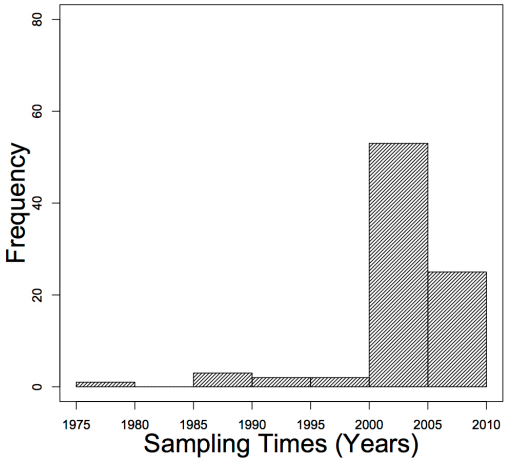
Appendix Figure 2. Histograms of sequence sampling times for each gene region.



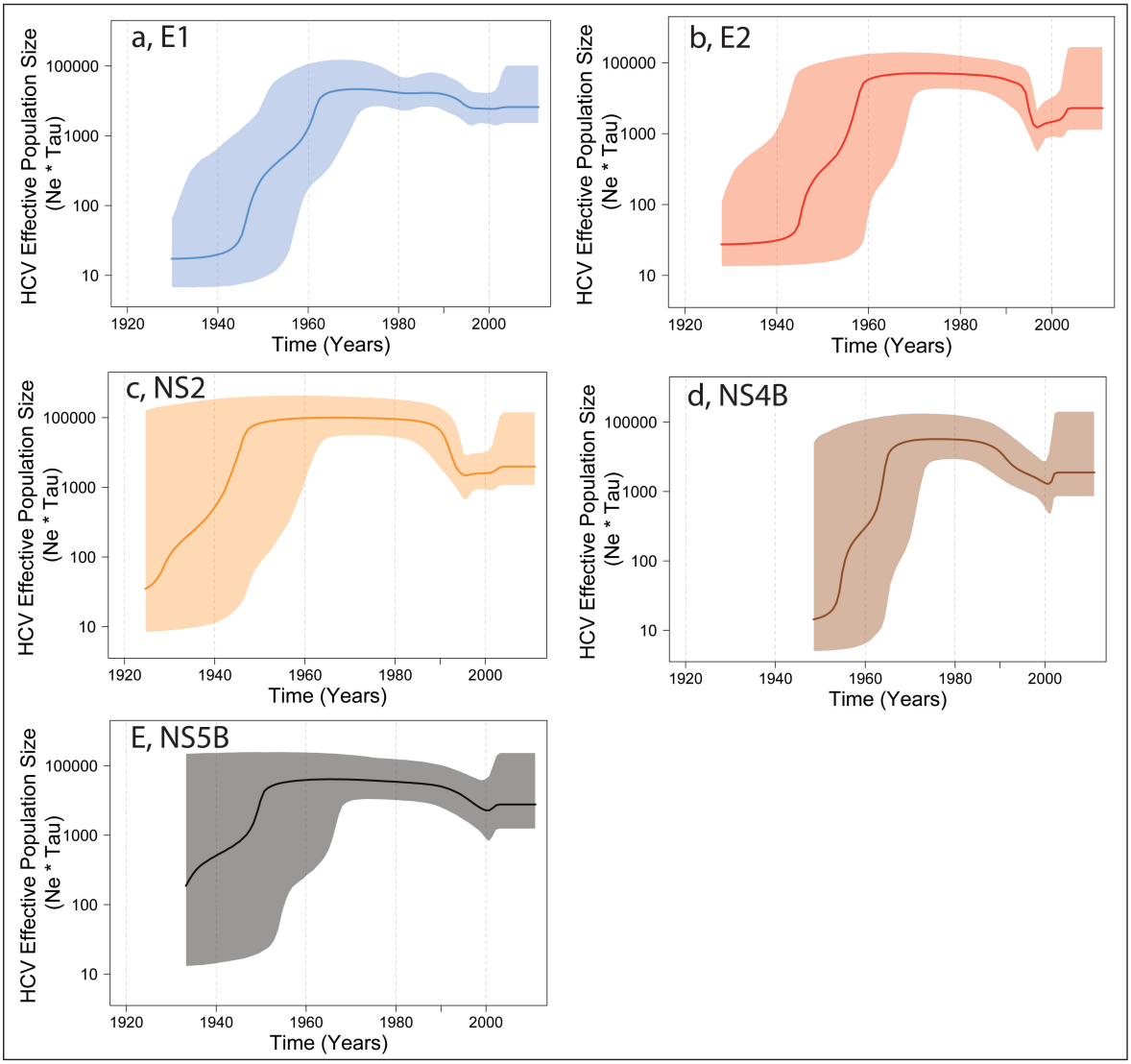
Appendix Figure 3. Time-scaled phylogenetic trees estimated for sequences sampled in North America for each HCV genotype 1a region analyzed.



Appendix Figure 4. Histogram of sequence sampling times from Magiorkinis et al.[4]



Appendix Figure 5. HCV effective number of infections estimated for each gene region. Coloured lines for each gene region represent the median estimated HCV effective number of infections (effective number of infections * generation time; Ne * Tau) through time estimated under a relaxed-clock model. The 95% highest probability density (HPD) for each gene region is shaded in colour. Areas of darker grey represent areas with the most HPD overlap across gene regions. Note the y-axis is on a log-scale units unaffected.



References

1. Talevich E, Invergo BM, Cock PJA, Chapman BA. Bio. Phylo: A unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. BMC bioinformatics 2012;13:209.
2. Gotoh O. An improved algorithm for matching biological sequences. Journal of molecular biology 1982;162:705-8.
3. Pond SLK, Frost SDW. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 2005;21:2531-3.
4. Magiorkinis G, Magiorkinis E, Paraskevis D, et al. The Global Spread of Hepatitis C Virus 1a and 1b: A Phylodynamic and Phylogeographic Analysis. PLoS Med 2009;6:e1000198.
5. Katoh K, Asimenos G, Toh H. Multiple alignment of DNA sequences with MAFFT. Methods in Molecular Biology 2009;537:39-64.
6. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE 2010;5:e9490.
7. Paradis E. Analysis of phylogenetics and evolution with R. New York: Springer; 2006.
8. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Molecular biology and evolution 2005;22:1185-92.
9. Frost SDW, Volz EM. Modelling tree shape and structure in viral phylodynamics. Philosophical Transactions of the Royal Society Series B 2013;368:1614.
10. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Molecular biology and evolution 2012;29:2157-67.
11. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 2012.